LEARN Journal: Language Education and Acquisition Research Network
ISSN: 2630-0672 (Print) | ISSN: 2672-9431 (Online)
Volume: 17, No: 2, July – December 2024

Language Institute, Thammasat University
https://so04.tci-thaijo.org/index.php/LEARN/index

Frequency Analysis, Distribution, and Coverage of Academic Words in Materials Science Research Articles: A Corpus-Based Study

Chariya Prapobratanakul

chariya.p@chula.ac.th, Chulalongkorn University Language Institute, Thailand
*Corresponding author, chariya.p@chula.ac.th

APA Citation:
Prapobratanakul, C. (2024). Frequency analysis, distribution, and coverage of academic words in materials science research articles: A corpus-based study. *LEARN Journal: Language Education and Acquisition Research Network, 17*(2), 793-813.

Received 14/04/2024

Received in revised form 20/06/2024

Accepted 08/07/2024

ABSTRACT

This study explored words in materials science research articles using New Academic Word List (NAWL) and New General Service List (NGSL) created by Browne at al. (2013) as reference lists. The objectives of this research were to explore the frequency and coverage of the NAWL and the NGSL in materials science research articles and identify frequently occurring academic words in materials science research articles that were not included in the NAWL and the NGSL. The corpus used for the analysis was compiled from 115 research articles published in five international journals related to the field of materials science. The findings revealed that the coverage of the NAWL words in the corpus of the present study was at 6.18 percent, together with the NGSL, the cumulative coverage of both word lists in the corpus of the present study was at 75.97 percent, and that there were 356 content words which did not appear in both the NAWL and the NGSL but did occur frequently in the corpus of materials science research articles. The pedagogical implications of these findings are discussed in relation to teaching academic

vocabulary to EFL learners of materials science who read and publish academic articles in English.

Keywords: academic vocabulary, word list coverage, materials science research articles

Introduction

Reading and writing academic texts in English pose notable challenges to students and researchers, particularly within contexts where English is a foreign language (EFL) (Evans & Morrison, 2018; Hyland, 2016; Lin & Morrison, 2021). An underlying issue contributing to this challenge is related to the specialized vocabulary used in academic texts (Evans & Morrison, 2018). In light of this concern, researchers have compiled academic corpora and developed word lists containing frequently encountered vocabulary in academic writings (e.g., Browne et al., 2013; Coxhead, 2000). These word lists are recognized as beneficial in assisting non-native English students and researchers regarding the vocabulary essential for comprehending and producing academic discourse (Coxhead & Nation, 2001).

The Academic Word List (AWL) (Coxhead, 2000) and the New Academic Word List (NAWL) (Browne et al. 2013) are examples of core academic word lists, focusing on words that frequently appeared in academic texts of multi-disciplinary corpora. One of their primary objectives was to provide learners of academic English across various fields with accessible vocabulary resources. Nonetheless, the concept of core academic word lists was questioned by several researchers, who argued that the selected words might be too general for a specific field of study, where the required vocabulary could differ from that of the core academic word list (e.g., Chen & Ge, 2007; Hyland & Tse, 2007). In response to these concerns, a number of studies have investigated the coverage of the existing core academic word lists in corpora specific to a discipline. They also identified words that frequently appear in academic texts of each discipline, which might vary in frequency compared to the core academic word lists. These specialized academic word lists are believed to offer useful vocabulary resources to EFL students or researchers who read or write academic texts in English within their discipline (e.g., Bunyarat, 2020; Chanasattru & Tangkiengsirisin, 2016; Gilmore & Millar, 2018; Ibrahim et al., 2014; Shabani & Tazik, 2014; Xodabande & Xodabande, 2020).

Regarding the core academic word lists, the New Academic Word List (NAWL) compiled by Browne et al. (2013) is one of the most recently

developed lists. The NAWL is a more recent development compared to the AWL, and it incorporates more contemporary academic vocabulary and additional academic words that have become increasingly important in contemporary academic discourse when compared to the AWL. The list consists of 967 words derived from a 288-million-word multi-disciplinary corpus of texts students are most likely to encounter in academic contexts. The disciplines covered by the NAWL were also broader and more diverse than those of the AWL. The NAWL, together with the New General Service List (NGSL), a general vocabulary word list developed by Browne et al. (2013), are reported to cover almost 92 per cent of the academic corpus used.

The extent of coverage provided by the NAWL in conjunction with the NGSL has been explored in academic corpora of specific disciplines. For instance, Gilmore and Millar (2018) compiled and analyzed the Specialized Corpus of Civil Engineering Research Articles (SCCERA). The findings revealed frequently encountered words in the corpus that were absent from the NAWL and the NGSL. These words were specifically linked to civil engineering research articles and were particularly beneficial to those who read and publish research articles in the field.

In Thailand, there is an increasing demand for post-graduate students, researchers, and academic staff to publish research articles in English and in international journals for graduation requirements, career advancement, or financial support (Khamkhien, 2016; Phothongsunan, 2016). Students and professionals in the field of materials science are no exception. Callister and Rethwisch (2018) defined materials science as an interdisciplinary field that "involves investigating the relationships that exist between the structures and the properties of materials" (p. 3). Carter and Paul (1991) asserted that materials science combines principles of physics, chemistry, engineering, and sometimes biology to understand the properties of materials. These materials can range from metals and ceramics to polymers and composites. Research in materials science is crucial for developing new materials or improving existing ones.

To support this field, English for Academic Purposes (EAP) and English for Specific Purposes (ESP) play a significant role in preparing students to read and write English academic texts. Given the students' diverse backgrounds and the limited time available in courses, one effective approach is to provide students with vocabulary that is commonly used and representative of their field. To this end, this study aims to investigate the frequency and the coverage of the NAWL and the NGSL words in the materials science research articles and identify frequently occurring academic words in materials science research articles that are not included in the NAWL and NGSL.

The findings of this study could help instructors of EFL students in the field of materials science in the selection of vocabulary to include in their courses or teaching materials, and non-native English speakers who read and write research articles in the field of materials science in expanding their knowledge for reading and writing research articles related to the materials science field.

Literature Review

The New Academic Word List (NAWL) and the New General Service List (NGSL)

The New Academic Word List (NAWL) was developed by Browne et al. in 2013. The list comprises 967 words which frequently appear in academic texts. These words were derived from a 288-million-word academic corpus, consisting of academic journals, non-fiction, student essays, academic discourse, and best-selling academic textbooks. The selection process excluded words present in the New General Service List (NGSL), also developed by Browne et al. (2013b). The NGSL is a list of 2809 most important high-frequency words in English that appeared in a 273-million-word sample of the Cambridge English Corpus (CEC).

The combined coverage of the NAWL and the NGSL was reported to be approximately 92 percent of the corpus used. This coverage rate was higher than the reported coverage rate of the Academic Word List (Coxhead, 2000) and the General Service List (West, 1953) of the same corpus. The NAWL and the NGSL are more recent developments compared to the AWL and the GSL. The NAWL and the NGSL incorporate more contemporary vocabulary and additional words that have become increasingly important in contemporary discourse when compared to the AWL and the GSL.

Regarding the methods of word list development, Laosrirattanachai and Laosrirattanachai (2021) concluded that the common criteria for constructing word lists include frequency, range, keyword analysis, lexical profiling, and expert verification. Frequency generally involves the inclusion of high-frequency words, while range considers the distribution of words across multiple sources. Keyword analysis identifies words with unusually high frequency in a target corpus compared to a reference corpus. Lexical profiling categorizes words into groups using reference lists. Lastly, expert verification involves experienced people validating a word list. Some word lists, such as the AWL, use more than one of these criteria to ensure that the selected words are not only common but also widely used across various texts.

Previous Studies on Coverage of Academic Words in Academic Research Articles

The analysis of academic words in academic research articles has been of particular interest during the past decades, partly driven by the increasing demand for English for Specific Purposes and English for Academic Purposes. Consequently, several academic word lists have been developed, and several studies were subsequently conducted to explore the effectiveness of these available core academic word lists in preparing EFL students for academic English discourse.

One of the most widely researched academic word lists is Coxhead's AWL. Several studies have been conducted to investigate the coverage rates of this list in academic texts of various specific disciplines and words which frequently appeared in these disciplines but were not included in the AWL (e.g., Bunyarat, 2020; Chanasattru & Tangkiengsirisin, 2016; Gilmore & Millar, 2018; Ibrahim, et al., 2018; Mozaffari & Moini, 2014; Shabani & Tazik, 2014; Xodabande & Xodabande, 2020). The analysis revealed that the vocabulary that frequently appeared in academic texts of these specific disciplines sometimes differed from the vocabulary included in the core academic word list. As a result, several discipline-specific academic word lists were created such as the Social Science Word List (SSWL) (Chanasattru & Tangkiengsirisin, 2016) and the Industrial Word List (IWL) (Bunyarat, 2020).

The more recent core academic word list is the NAWL, created by Browne et al. in 2013. The list was developed to represent academic texts of various disciplines similar to the AWL. There has been, however, limited research exploring the coverage of the NAWL in academic research articles, particularly on specific disciplines, which might be partly due to its recent development compared to the AWL. One such study was conducted by Gilmore and Millar (2018), focusing on identifying vocabulary associated with the field of Civil Engineering using a corpus-based investigation of the 8-million-word Specialized Corpus of Civil Engineering Research Articles (SCCERA). They compared corpus-derived keywords with the NGSL and NAWL to identify the commonly occurring words in general and academic English, and those which did not occur in either word list but were specific to Civil Engineering. The findings indicated that the NGSL covered approximately 60.4 percent of the entire corpus, while the NAWL covered around 16.9 percent. There were, however, 22.8 percent which was neither in the NGSL nor the NAWL. The researchers concluded that the words in this last category were specific to the academic texts in the field of Civil Engineering.

The coverage of the NAWL has also been explored in other contexts. Nonetheless, there have been no studies of the NAWL in materials science

academic research articles. The frequency analysis of the corpus of materials science academic research articles with the NAWL as a reference list would offer useful resources for both students and researchers in this field.

Research Questions

1. What are the frequency and the coverage of the NAWL and the NGSL words in the materials science research articles?
2. What are the frequently occurring academic words in materials science research articles that are not included in the NAWL and NGSL?

Methodology

Data collection

Corpus compilation

For this study, a corpus of research articles in materials science was specifically compiled from 115 open-access research articles published in five international journals related to the field of materials science, namely Nature Energy, Nature Materials, Nature Nanotechnology, Nature Photonics and Nature Electronics. The journals were selected regarding the impact factors based on data from 2022 to ensure their significance in the materials science research community. However, the highest-ranked journal, the Nature Reviews Materials journal was excluded from the study because the journal mainly published review articles and did not feature any research articles in 2023. All open-access research articles published in 2023 are compiled for the corpus of this study to avoid using outdated terminology and to ensure the use of up-to-date vocabulary in contemporary materials science research. The composition of the corpus and the impact factor of each journal selected for this study are presented in Table 1.

Table 1

Composition of the Materials Science Research Articles Corpus

	Journal name (Impact Factor, 2022)	No. articles	Tokens	Types
1	Nature Energy (19.588)	17	110985	8870
2	Nature Materials (13.874)	25	153991	13736
3	Nature Nanotechnology (13.141)	26	172012	15568
4	Nature Photonics (11.774)	31	165743	12770
5	Nature Electronics (10.927)	16	93815	9908
	Total	115	696546	60852

Note. Data for the impact factors from SCImago (2023)

These selected journals published research articles focusing on different subfields within materials science. To illustrate, the Nature Energy journal focused on energy-related research, including research in energy materials. The Nature Materials journal covered a broad spectrum of material-related research, including studies on diverse material types and applications. The Nature Nanotechnology journal focused on research at the nanoscale, which is a rapidly growing area within materials science. The Nature Photonics journal covered research in the field of photonics, which has significant implications for materials science. Lastly, the Nature Electronics journal published research articles in the field of electronics, including electronics materials.

The research articles were collected in their electronic versions with their graphs, charts, diagrams, equations, reference lists, appendices, footnotes, and acknowledgements removed. The units of analysis were word tokens (i.e., the number of occurrences of each type) and word types (i.e., single word forms) (Bauer & Nation, 1993). A sub-corpus of each journal was then created.

The NAWL and the NGSL

The NAWL and the NGSL (Version 1.2) compiled by Browne et al. (2013a; 2013b) were employed in this study. The NAWL and the NGSL (Version 1.2) consisted of 2809 and 957 words, respectively. The lists were downloaded from the New General Service List website.

Lexical Analysis Software

In this study, the Antconc (Version 4.2.4) software program (Anthony, 2023) was employed to analyze word frequencies. The Stop List function of the AntConc program was applied to sort out the function words using O'Shea's function word lists (O'Shea, 2024). The Compleat Lexical Tutor (Version 4.0) software tool (Cobb, 2023) was used to find the coverage of the reference word lists and the high-frequency non-NAWL words in the corpus of materials science research articles.

Data Analysis

To investigate the frequency and coverage of the NAWL and NGSL words in the materials science research articles, this study initially determined the frequency and distribution of words in the corpus of materials science research articles. Then, using the NAWL and NGSL as reference lists, this

study identified academic and general words present in the corpus and analyzed their coverage.

To identify frequently occurring academic words in materials science research articles that were not included in the NAWL and NGSL, this study used the NAWL and NGSL as reference lists to filter out the vocabulary in the corpus, excluding words present in these word lists. The researcher then manually sorted high-frequency content words from the corpus of materials science research articles that were not included in the NAWL and NGSL, by applying Coxhead's word selection criteria (Coxhead, 2000). Coxhead used these criteria as principles of selection for academic words to be included in her word list. These criteria were widely used tool in academic vocabulary research. Her method included multiple criteria, i.e., range, frequency, and uniformity of frequency, ensuring that the selected words were not only common but also widely used across various texts.

The details of Coxhead's word selection criteria applied in this study are as follows.

1. Range: Coxhead compiled her corpus by collecting language data from 28 different sources. She asserted that words appearing in more than half of her sources met the range criterion. Therefore, for this study, words that appeared in at least three out of the five journals were considered to meet this criterion.

2. Frequency: In her study, Coxhead created a corpus of 3,500,000 tokens, and any word occurring at least 100 times was considered to meet the frequency criterion. This criterion was also used by Laosrirattanachai and Ruangjaroon (2021) in their study, where they proposed the following equation to determine the minimum frequency required for a word to meet this criterion.

$$100 \times \frac{\text{total word tokens in a corpus}}{3{,}500{,}000}$$

The present study applied this equation as follows:

$$100 \times \frac{696{,}546}{3{,}500{,}000}$$

From the equation, for a word to meet this criterion, it had to occur at least 19.90 times in the corpus of materials science research articles. Therefore, in this study, words appearing at least twenty times in the corpus of materials science research articles would pass this criterion.

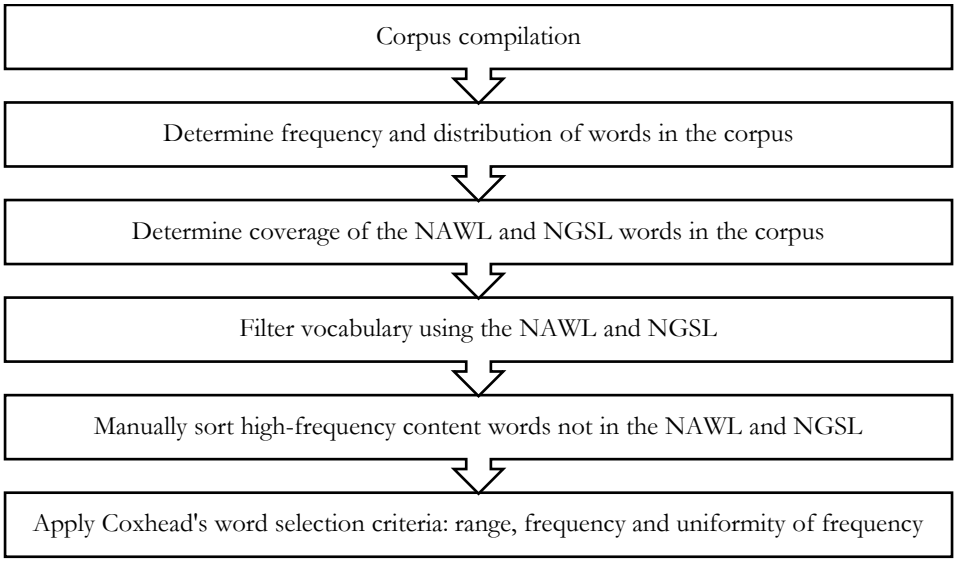
3. Uniformity of frequency: In Coxhead's study, the average number of words per section in her corpus was 875,000, with a minimum occurrence

threshold of 10 times for a word to be included in her study. Therefore, the minimum percentage for inclusion could be calculated as 10/875,000 (= 0.0011%). Applying this formula to the current study resulted in a minimum occurrence for inclusion of 10/875,000 x (696,546/5) = 1.59 words. Thus, in this study, words appearing at least twice in each journal of the corpus of materials science research articles would pass this criterion.

Content words that passed these three criteria would be classified as high-frequency academic off-list words in the corpus of materials science research articles. The detailed steps for data analysis of this study are presented in Figure 1.

Figure 1

Flowchart of Steps in Data Analysis



This flowchart presents the sequential steps and criteria used in the study to analyze the corpus and identify academic vocabulary not included in the NAWL and the NGSL.

Results

Distribution and Coverage of The NAWL and the NGSL Words In The Corpus Of Materials Science Research Articles

Table 2 presents the distribution and coverage of both the NGSL and the NAWL words in the entire corpus of materials science research articles.

Table 2

The Distribution and Coverage of The NGSL And NAWL Words in the Corpus of Materials Science Research Articles

	Coverage	
	Tokens	**Percentage**
NGSL Level 1	486102	69.79
NAWL	43041	6.18
Off-list	167403	24.03
Total	696546	100

The NGSL words accounted for a total coverage of 486,102 tokens, equivalent to 69.79 percent of the entire corpus. The NAWL accounted for 43,041 tokens, representing 6.18 percent of the entire corpus. Therefore, the combined coverage offered by the NGSL and NAWL was 75.97 percent, indicating that 24.03 percent or 167,403 tokens were beyond the coverage of these lists.

The details of the coverage rate of NAWL words in each materials science journal are shown in Table 3.

Table 3

Coverage of NAWL Words in the Corpus of Materials Science Research Articles

Journals	**Running words**	**NAWL word forms**	**% of NAWL coverage**
Nature Energy	110985	5806	5.23
Nature Materials	153991	9818	6.38
Nature Nanotechnology	172012	10117	5.88
Nature Photonics	165743	11485	6.93
Nature Electronics	93815	5815	6.20

The results revealed that Nature Photonics exhibited the highest coverage of the NAWL words at 6.93 percent, followed by Nature Materials at 6.38 percent, Nature Electronics at 6.20 percent, and Nature Nanotechnology at 6.20 percent, respectively. The journal with the lowest percentage of NAWL coverage was Nature Energy at 5.23 percent.

Pertaining to the frequently occurring NAWL words in the corpus of materials science research articles, among the 957 NAWL words, 659 words (68.86%) occurred in the corpus. There were 298 words (31.14%) that did not appear at all. The top 50 NAWL words and their frequency in the Materials Science Research Articles Corpus are shown in Table 4.

Table 4

Top 50 NAWL Words in the Materials Science Research Articles Corpus

Rank	Words	Freq.	Rank	Words	Freq.
1	optical	1081	26	probe	215
2	quantum	677	27	bulk	197
3	electron	580	28	matrix	195
4	magnetic	552	29	simulation	194
5	detection	422	30	characterization	192
6	transmission	407	31	carrier	191
7	intensity	398	32	correlation	191
8	thermal	394	33	atomic	190
9	beam	368	34	absorption	189
10	crystal	360	35	array	188
11	wavelength	344	36	nonlinear	184
12	emission	319	37	dependence	180
13	distribution	318	38	grid	178
14	interface	307	39	membrane	168
15	ray	281	40	integration	165
16	dimensional	280	41	configuration	160
17	linear	277	42	obtain	159
18	spectrum	276	43	diameter	157
19	molecular	271	44	mechanical	157
20	axis	262	45	impact	154
21	ion	259	46	magnitude	154
22	spatial	247	47	threshold	153
23	domain	244	48	temporal	150
24	fiber	242	49	transformation	150
25	amplitude	218	50	feedback	146

As shown in Table 4, the word "optical" appeared as the most frequently used NAWL word in the corpus of the current study, occurring 1081 times. Other NAWL word items such as "quantum," "electron," "magnetic," "detection," and "transmission" were also used frequently, appearing over 400 times. Several NAWL words which occurred frequently in the corpus of the present study, e.g., "distribution," "obtain," "impact," "domain," "ion," or "matrix", which ranked in the top 20 in the corpus of Browne et al. (2013), were also similarly used very often in the corpus of the current study. Nonetheless, it is notable that certain words which were frequent in the corpus of Browne et al. (2013), such as "repertoire," "semantic," "cognitive," or "linguistic" did not appear as frequently in the corpus of the present study.

Non-NAWL Content Words in The Corpus of Materials Science Research Articles

From the analysis of frequently occurring academic words in the corpus of materials science research articles that were not included in the NAWL, the data revealed that 356 off-list word types meet the criteria for high-frequency words in the current study: appearing in at least three out of five journals, 20 times in the corpus, and twice in each of the five journals. These 356 off-list word types and their frequency in the corpus are shown in the appendix. The most frequently occurring off-list word types in the corpus were "supplementary," "laser," "photon," "voltage," and "graphene." The word "supplementary" appeared 2508 times, more than the other word types, as the authors used this word to refer to content provided at the end of articles that was not part of the main text, such as "supplementary note," "supplementary table," or "supplementary section." The words "laser," "photon," "voltage," and "graphene" appeared more than 400 times in the corpus. Word types that occurred least frequently at 20 times in the off-list list were "epitaxy," "evenly," "traverse," "tensor," "medians," and "microspheres."

Table 6 presents the percentage coverage provided by the NGSL, NAWL, and the 356 high-frequency off-list words for each journal.

Table 6

Percentage Coverage Provided by the NGSL, NAWL, And The 356 High-Frequency Off-List Words for Each Journal

Journal	% coverage			Total % coverage
	NGSL	NAWL	356 off-list words	
Nature Energy	73.04	5.23	2.70	80.97
Nature Materials	65.81	6.38	4.36	76.55
Nature Nanotechnology	64.51	5.88	3.64	74.03
Nature Photonics	69.64	6.93	4.96	81.53
Nature Electronics	69.82	6.20	5.42	81.44

Nature Electronics showed the highest coverage of the 356 off-list at 5.42 percent. This was followed by Nature Photonics, Nature Materials, and Nature Nanotechnology at 4.96 percent, 4.36 percent, and 3.64 percent, respectively. The Nature Energy journal exhibited the lowest coverage of the NAWL words at 2.70 percent. The coverage provided by the NGSL, NAWL,

and the 356 off-list words ranged from 74.03 to 81.53 percent in the five journals.

Discussion

The results showed that the combined coverage percentage of the NGSL and NAWL in the corpus of materials science research articles, at 75.97 percent, was lower than the 92% coverage of these word lists reported by Browne et al. (2013) for their corpus. This difference might stem from the types of text included in the corpus used for the NAWL and NGSL compared to that for the present study. The NAWL was derived from a combination of both spoken and written academic corpora and included diverse text types from multiple academic disciplines including academic journals, non-fiction student essays, and academic textbooks. However, the corpus for the present study was compiled solely from materials science research articles, representing only the written genre.

However, the data showed that the coverage rate of NAWL words in the present study did not much differ from Gilmore and Millar's (2018) findings. In their study, the coverage percentage of NAWL words ranged from 3.7 to 6.3 percent, which exhibited minor differences compared to the coverage range of 5.23 to 6.93 percent found in the present study. These findings might be due to the fact that the genre of the texts in both Gilmore and Millar's (2018) corpus and the corpus of the present study was the same namely academic writing and research articles.

With regard to the most frequently used off-list words, as can be seen from the list in the appendix, several of these high-frequency non-NAWL words are words related to the properties of materials, such as "polymer," "crystalline," "annealing," "conductivity," "ionic," "lattice," "metallic," "nucleation," "oxide," and "propagation". The findings of this study provide evidence that specialized academic corpora exhibit some high-frequency words that differ from those found in academic corpora which contain texts from various fields of study and from various genres. These findings are also in line with previous studies conducted with academic specialized corpora. Gilmore and Millar's (2018) study, which examined keywords in their Specialized Corpus of Civil Engineering Research Articles, found that keywords in their corpus that were absent from the NGSL and NAWL were words typical of civil engineering discourse. Similarly, a study by Vongpumivitch et al. (2009) revealed that vocabulary related to the field of applied linguistics that was not included in the AWL appeared frequently in their Applied Linguistics Research Articles Corpus.

Regarding the coverage provided by the NGSL, NAWL, and the 356 off-list words, it ranged from 74.03 to 81.53 percent in the five journals. This

coverage rate still falls below the estimated comprehension threshold, which is generally reported to be between 95 and 98 percent (Laufer, 1989; Nation, 2006). Nevertheless, despite the relatively low coverage percentages ranging from approximately 2.70 to 5.42 percent, the 356 off-list words identified in this study might be beneficial as supplementary vocabulary to the existing word lists for EFL learners or non-native English readers and writers of research articles in materials science who would like to learn and familiarize themselves with high-frequency words found in the research articles in materials science journals.

Limitations and Recommendations

The corpus of materials science research articles used in this study comprises approximately 700,000 words. This is relatively small compared to the 288-million-word academic corpus used by Browne et al. (2013) for the NAWL, and the 8-million-word Specialized Corpus of Civil Engineering Research Articles used by Gilmore and Millar (2018). The limited size is partly due to the criteria for the selection of articles from 2023 to ensure contemporary vocabulary, and also from only the top five high-impact journals to guarantee high-quality research. While this focused selection provides a current dataset, it may not cover all sub-disciplines comprehensively. The articles included in the corpus, therefore, might not be fully representative of the entire field. Thus, it is recommended that future studies increase the corpus size and include a broader range of journals and sub-disciplines within materials science to enhance the comprehensiveness and representativeness of the corpus.

Another limitation is that this study specifically examined the frequency analysis of academic words in the corpus of materials science research articles. Hence, further investigations into how these words are used in context are suggested, as they would offer valuable insights into their usage patterns. Such insights would serve as beneficial resources for learners and practitioners who read and write research articles in materials science.

The last limitation is that this study explored only high-frequency words. Future research examining mid-frequency words in the corpus of materials science research articles is also recommended. Investigations into mid-frequency words would offer additional valuable resources for EFL learners and non-native English researchers in the materials science field, enabling them to increase necessary vocabulary knowledge.

Conclusion and Implications

The present study investigated the coverage of the NAWL words in the corpus of materials science research articles from five major research journals in materials science. The findings revealed that the coverage of the NAWL words in the corpus of the present study was at 6.18 percent, and together with the NGSL, the cumulative coverage of both word lists in the corpus of the present study was at 75.97 percent. This percentage was lower than the 92 percent coverage of the academic corpus reported by Browne et al. (2013). In addition to coverage rates, this study examined frequently used academic words in the corpus of materials science research articles not included in the NAWL and the NGSL. Following Coxhead's word selection criteria (Coxhead, 2000), 356 off-list content words were identified as occurring at high frequency in the corpus of materials science research articles.

The findings of this study, namely the high-frequency NAWL words appeared in the corpus of materials science research articles and the 356 off-list content words, could serve as resources for instructors of EFL students in the field of materials science, aiding in the selection of vocabulary to include in their courses or teaching materials. Moreover, learners of academic English and non-native English speakers who read and write research articles in the field of materials science could also benefit from the lists of high-frequency NAWL words and frequently occurred content words that do not appear in the NAWL and the NGSL presented in this paper. These lists could expand their vocabulary knowledge for reading and writing research articles related to the materials science field.

Regarding teaching academic vocabulary to EFL students, teachers can integrate the high-frequency academic vocabulary identified in this study into their lessons. These academic words can be directly introduced in a language classroom whose aim is for students to read and write research articles related to materials science. To familiarize students with this vocabulary, teachers might provide students with materials science research articles and ask students to identify and highlight the high-frequency words from the lists provided in this paper. Students and teachers can then discuss how these words are used in context. This activity could help students recognize academic vocabulary in research articles and understand its application. Another suggested activity is asking students to write a summary of a materials science research article using a specified number of high-frequency words provided in this paper. After that students and teachers could discuss the summaries in class, focusing on the correct and effective use of vocabulary. This could enhance both students' reading comprehension

skills and their ability to use academic vocabulary in writing materials science research articles.

In conclusion, this research provides empirical data on the high-frequency words appeared in materials science research articles. The findings may encourage further studies and applications in other specialized fields, thereby enhancing the overall effectiveness of ESP programs.

About the Author

Chariya Prapobratanakul: A lecturer at Chulalongkorn University Language Institute, Thailand. She holds a Ph.D. in English as an International Language from Chulalongkorn University, Thailand. Her research interests include Second Language Acquisition, and English for Specific and Academic Purposes.

References

Anthony, L. (2023). *AntConc* (Version 4.2.4) [Computer Software]. Tokyo, Japan: Waseda University. https://www.laurenceanthony.net/software

Bauer, L. & Nation, P. (1993). Word families. *International Journal of Lexicography, 6*(4), 253-279. https://doi.org/10.1093/ijl/6.4.253

Browne, C., Culligan, B. & Phillips, J. (2013). The New Academic Word List. Retrieved from http://www.newgeneralservicelist.org.

Browne, C., Culligan, B. & Phillips, J. (2013). The New General Service List. Retrieved from http://www.newgeneralservicelist.org.

Bunyarat, P. (2020). A corpus-based study of specialized vocabularies in industrial work for developing industrial word lists of English for science and technology class. *Journal of Language and Culture, 39*(2), 122-151. https://so03.tci-thaijo.org/index.php/JLC/article/view/248689

Callister, W. D. & Rethwisch D. G. (2018). Materials science and engineering: An introduction. 10th Ed. Wiley.

Carter, G. F. & Paul D. E. (1991). Materials Science and Engineering. ASM International: Materials Park, Ohio.

Chanasattru, S. & Tangkiengsirisin, S. (2016). Developing of a high frequency word list in social sciences. *Journal of English Studies, 11*, 41-87. https://so04.tci-thaijo.org/index.php/jsel/article/view/73309

Chen, Q. & Ge, G. (2007). A corpus-base lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles (RAs). *English for Specific Purposes, 26*, 502-514. https://doi.org/10.1016/j.esp.2007.04.003

Cobb, T. *Compleat Lexical Tutor* (Version 4.0) [Computer Software]. https://www.lextutor.ca/cgi-bin/tl_compare/

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238. https://doi.org/10.2307/3587951

Evans, S. & Morrison, B. (2018). Adjusting to higher education in Hong Kong: the influence of school medium of instruction. *International Journal of Bilingual Education and Bilingualism, 21*(8), 1016-1029. https://doi.org/10.1080/13670050.2016.1228600

Gilmore, A., & Millar, N. (2018). The language of civil engineering research articles: A corpus-based approach. *English for Specific Purposes, 51*, 1-17. https://doi.org/10.1016/j.esp.2018.02.002

Hyland, F. (2016). Challenges faced by second language doctoral student writers in Hong Kong and their writing strategies. *Australian Review of Applied Linguistics, 39*(2), 158-180. http://dx.doi.org/10.1075/aral.39.2.04hyl

Hyland, K. & Tse, P. (2007). Is there an "academic vocabulary"? *TESOL Quaterly, 41*(2), 235-253. https://www.jstor.org/stable/40264352

Ibrahim, E., Shah, M., Abudukeremu, M. (2018). A corpus-based lexical study of the frequency, coverage and distribution of academic vocabulary in Islamic academic research articles. *The Journal of Social Sciences Research*, *Special Issue 2*, 570-577. https://ideas.repec.org/a/arp/tjssrr/2018p570-577.html

Khamkhien, A. (2016). Structural patterns and linguistic features of research articles on applied linguistics: Evidence from introduction sections to methods section. *Journal of Linguistics and Language Teaching, 7*, 11-37. https://linguisticsandlanguageteaching.blogspot.com/2016/07/blog-post.html

Laosrirattanachai, P. & Laosrirattanachai, P. (2021). Applying lexical profiling to construct technical word lists for Thai tourist guides. *PASAA: A Journal of Language Teaching and Learning*, *62*(1), 61-91. https://www.culi.chula.ac.th/Images/asset/pasaa_journal/file-4-45-tluq36161081.pdf

Laosrirattanachai, P. & Ruangjaroon, S. (2021). Corpus-based creation of tourism, hotel and airline business word lists. *LEARN Journal, 14*(1), 50-86. https://so04.tcithaijo.org/index.php/LEARN/article/view/248677

Laufer, B. (1989). What percentage of lexis is essential for comprehension. In Lauren, C. and Nordman, M. (Eds), From humans thinking to thinking machines. Multilingual Matters (pp. 316-23).

Lin, L., Morrison, B. (2021). Challenges in academic writing: Perspectives of Engineering faculty and L2 postgraduate research students. *English for Specific Purposes, 63*, 59-70. https://doi.org/10.1016/j.esp.2021.03.004

Mozaffari, A. & Moini, R. (2014). Academic words in educational research articles: A corpus-based study. *Procedia – Social and Behavioral Sciences, 98*, 1290-1296. https://doi.org/10.1016/j.sbspro.2014.03.545

Nation, I. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review, 63*(1), 59-82. https://doi.org/10.3138/cmlr.63.1.59

O'Shea, J. (2024). Function word lists. https://semanticsimilarity. wordpress.com /function-word-lists/

Phothongsunan, S. (2016). Thai university academics' challenges of writing for publication in English. *Theory and Practice in Language Studies*, 6(4), 681-685. http://dx.doi.org/10.17507/tpls.0604.04

SCImago, (2023). SJR — SCImago Journal & Country Rank [Portal]. https://www.scimagojr.com

Shabani, M. B. & Tazik, K. (2014). Coxhead's AWL across ESP and Asian EFL journal research articles (RAs): A corpus-based lexical study. *Procedia – Social and Behavioral Sciences, 98,* 1722-1728. https://doi.org/10.1016/j.sbspro.2014.03.599

Vongpumivitch, V., Huang, J. & Chang, Y. (2009). Frequency analysis of the words in the Academic Word List (AWL) and non-AWL content words in applied linguistics research papers. *English for Specific Purposes, 28*, 33-41. https://doi.org/10.1016/j.esp.2008.08.003

Xodabande, I. & Xodabande, N. (2020). Academic vocabulary in psychology research articles: A corpus-based study. *MAXTESOL Journal, 44*(3), 1-21. https://files.eric.ed.gov/fulltext/EJ1310955.pdf

West, M. P. (1953). A general service list of English words: With semantic frequencies and a supplementary word-list for the writing of popular science and technology. Longman Green.

Appendix
356 off-list word types in the corpus of materials science research articles

Words	Freq.	Words	Freq.	Words	Freq.
supplementary	2508	emitter	84	stabilization	44
laser	726	ensemble	84	epitaxial	43
photon	603	substantially	84	favourable	43
voltage	551	robust	83	flakes	43
graphene	425	calibration	82	polarity	43
electrode	360	optically	82	quartz	43
bandwidth	331	ambient	81	theoretically	43
electrodes	323	adhesion	80	couplings	42
electrical	316	degradation	80	desirable	42
acoustic	307	focal	80	encapsulation	42
buffer	302	enhancement	79	hamiltonian	42

silicon	302	linearly	79	hundreds	42
fluorescence	293	negligible	79	mismatch	42
resonance	289	circuits	78	purity	42
spectral	288	illumination	78	suppression	42
polarization	262	scalable	78	validate	42
excitation	246	etching	77	boron	41
inset	236	passive	77	mechanically	41
microscopy	235	programmable	77	methyl	41
conversion	234	receiver	77	sputtering	41
fabrication	233	annealed	76	counterparts	40
dispersion	230	baseline	76	reproducibility	40
diffraction	223	photocurrent	76	suppress	40
fabricated	223	aperture	75	tunability	40
width	215	aligned	74	uniformity	40
activation	203	bilayer	74	electromagnetic	39
schematic	201	linewidth	74	iodide	39
amplifier	198	ultraviolet	72	micrometre	39
modulation	195	diamond	71	rigid	39
vivo	194	grating	71	spatiotemporal	39
dashed	188	metallic	71	etched	38
resonant	186	dataset	70	heterogeneous	38
zero	184	pixels	70	intrinsically	38
tunnelling	183	analogue	69	optimize	38
normalized	181	circular	69	spherical	38
lattice	178	million	69	visualization	38
topological	176	transient	69	attenuation	37
symmetry	174	nitride	68	microscopic	37
microscope	172	spatially	68	diodes	36
amplification	170	compatible	67	isopropanol	36
crystalline	164	equipped	67	orbit	36
polymer	164	adsorption	66	piezo	36
semiconductor	164	multiplexing	66	rinsed	36
intrinsic	159	precision	66	selectively	36
sigma	158	extraction	65	topologically	36
doping	157	homogeneous	65	visualized	36
circuit	155	argon	64	absent	35
interference	154	insertion	64	accordingly	35
propagation	154	microwave	64	agilent	35
sensor	154	simultaneous	64	analogous	35
radiative	153	spectrometer	64	controller	35
detector	148	thermally	64	coulomb	35
spectroscopy	144	ultra	64	mitigate	35
optimized	143	calibrated	63	modulate	35
electrochemical	136	confinement	63	optimizing	35
excitons	135	ethanol	63	originating	35
experimentally	135	lasers	63	topology	35
pixel	134	deionized	62	voltammetry	35
deposition	131	diode	62	waveforms	35
oxide	130	inorganic	62	asymmetric	34
photonics	130	multiplication	62	functionalities	34

situ	129	optimization	62	fused	34
dots	128	encapsulated	61	indium	34
localized	128	dotted	59	proximity	34
hybrid	125	harmonic	59	sequential	34
resonances	123	polymers	59	analytical	33
fluorescent	120	proportional	59	biologically	33
geometry	120	convolution	58	nano	33
reciprocal	120	generator	58	ohmic	33
monolayer	118	mask	58	vertically	33
polarized	118	perpendicular	58	cubic	32
interlayer	117	insulating	57	fabricate	32
refractive	117	inverse	57	oscillator	32
tunable	117	purple	57	quenching	32
transducer	116	suppressed	57	rectangular	32
aldrich	115	vapour	56	deduced	31
amplified	113	acceptor	55	originate	31
orbital	112	capacitance	55	resultant	31
setup	110	fluctuations	55	anisotropic	30
readout	107	geometrics	55	controllable	30
synthesized	107	optoelectronic	55	integrity	30
heterostructure	106	quasi	55	originates	30
voltages	106	aqueous	54	python	30
additionally	104	considerably	54	ultrashort	30
infrared	103	reconfigurable	54	vibrations	30
conductivity	102	copper	53	adhesive	29
plasma	102	dual	53	healthcare	29
confocal	101	latency	53	nanometres	29
dipole	101	nucleation	52	reproducible	29
symmetric	101	superconducting	52	sequentially	29
transverse	101	adsorbed	51	addressable	28
bandgap	100	responsive	51	applicable	28
complementary	100	broadband	50	magneto	28
annealing	99	insets	50	modulating	28
dielectric	98	lithography	50	snapshots	28
assembled	97	verify	50	stochastic	28
nanoscale	97	magnification	49	bare	27
photoluminescenc	97	acetone	48	femtosecond	27
e	97	dispersive	48	illuminated	27
repetition	96	geometric	48	nanometre	27
alignment	96	lens	48	centimetre	26
heterostructures	96	notable	48	extinction	26
localization	95	operational	48	logarithmic	25
junction	94	histogram	47	successive	25
phonon	94	micrograph	47	tungsten	25
transparent	94	offset	47	noticeable	24
vacuum	93	triangular	47	solely	24
eight	93	verified	47	imaginary	23
overnight	93	continuously	46	melt	23
ralaxation	92	negatively	46	titanium	22
respective	91	sectional	46	wavevector	22

electro	91	waveform	46	chloroform	21
ionic	91	functionality	45	singularities	21
kinetics	91	phonons	45	epitaxy	20
modulated	89	computational	44	evenly	20
sensors	86	evaporation	44	traverse	20
angular	86	exfoliated	44	tensor	20
gaussian	85	precursors	44	medians	20
notably	84	roughness	44	microspheres	20